Complete Topological Mapping of a Cellular Protein Interactome Reveals Bow-Tie Motifs as Ubiquitous Connectors of Protein Complexes

Niss, Kristoffer; Gomez-Casado, Cristina; Hjaltelin, Jessica X; Joeris, Thorsten; Agace, William W.; Belling, Kirstine G; Brunak, Søren

Published in:
Cell Reports

DOI:
10.1016/j.celrep.2020.107763

Publication date:
2020

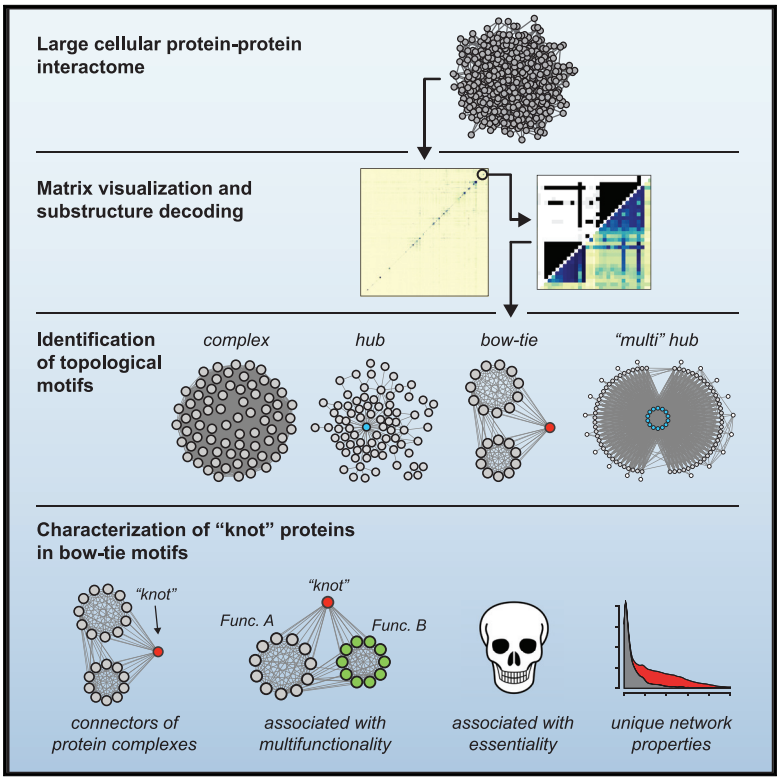
Document version
Publisher's PDF, also known as Version of record

Cell Reports

Article

Complete Topological Mapping of a Cellular Protein Interactome Reveals Bow-Tie Motifs as Ubiquitous Connectors of Protein Complexes

Graphical Abstract



Highlights

- Visual overview of the network topology of a cellular protein-protein interactome
- A scaffold of bow-tie motifs connects the majority of protein complexes
- Central proteins of bow-tie motifs, called "knots," have unique network properties
- "Knot" proteins are associated with multifunctionality and essentiality

Authors

Kristoffer Niss, Cristina Gomez-Casado, Jessica X. Hjaltelin, Thorsten Joeris, William W. Agace, Kirstine G. Belling, Søren Brunak

Correspondence

soren.brunak@cpr.ku.dk

In Brief

Niss et al. show that topological motifs called bow-ties create a scaffold within the cellular protein interactome that connects a majority of protein complexes. The central proteins of these motifs are found to be associated with multifunctionality and cellular essentiality, display unique network properties, and are expressed widely across cells and tissues.

Niss et al., 2020, Cell Reports *31*, 107763
June 16, 2020
https://doi.org/10.1016/j.celrep.2020.107763

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

Complete Topological Mapping of a Cellular Protein Interactome Reveals Bow-Tie Motifs as Ubiquitous Connectors of Protein Complexes

Kristoffer Niss,[1] Cristina Gomez-Casado,[2,3] Jessica X. Hjaltelin,[1] Thorsten Joeris,[2] William W. Agace,[2,4] Kirstine G. Belling,[1] and Søren Brunak[1,5,*]

[1]Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, 2200 Copenhagen, Denmark

[2]Immunology Section, Lund University, BMC D14, 221-84 Lund, Sweden

[3]Institute of Applied Molecular Medicine, Faculty of Medicine, San Pablo CEU University, 28925 Madrid, Spain

[4]Mucosal Immunology Group, Department of Health Technology, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark

[5]Lead Contact

*Correspondence: soren.brunak@cpr.ku.dk

https://doi.org/10.1016/j.celrep.2020.107763

SUMMARY

The network topology of a protein interactome is shaped by the function of each protein, making it a resource of functional knowledge in tissues and in single cells. Today, this resource is underused, as complete network topology characterization has proved difficult for large protein interactomes. We apply a matrix visualization and decoding approach to a physical protein interactome of a dendritic cell, thereby characterizing its topology with no prior assumptions of structure. We discover 294 proteins, each forming topological motifs called "bow-ties" that tie together the majority of observed protein complexes. The central proteins of these bow-ties have unique network properties, display multifunctional capabilities, are enriched for essential proteins, and are widely expressed in other cells and tissues. Collectively, the bow-tie motifs are a pervasive and previously unnoted topological trend in cellular interactomes. As such, these results provide fundamental knowledge on how intracellular protein connectivity is organized and operates.

INTRODUCTION

Protein-protein interactomes are large assemblies of protein-protein interactions (PPIs) (Licata et al., 2012; Chatr-Aryamontri et al., 2017; Li et al., 2017) and can be pruned such that they model the proteome of a specific tissue, cell type, or cellular state (Magger et al., 2012; Pedersen et al., 2017; Yeger-Lotem and Sharan, 2015). Because multiple algorithms can predict protein function on the basis of local network topology (Chua et al., 2006; Hishigaki et al., 2001; Zhao et al., 2016), it follows that the function of proteins must have shaped the network topology. This statement is also true for the function of protein assemblies. Protein assemblies form specific topological structures (motifs), whose layout indicates their collective function: proteins arranged in pathways often transduce signals, proteins with many PPIs frequently regulate their neighboring proteins (Lefebvre et al., 2010), and proteins in highly interconnected complexes often work in sync on the same biological process (Hu et al., 2016). Hence, network topology should be considered a resource of functional information. By applying network topology analysis, this resource can be mined.

However, complete characterization of the network topology of large protein interactomes is difficult. When interactomes exceed a couple of hundred proteins, graph visualization using edges and nodes becomes uninformative. This issue is known as the hairball problem. Previous studies have avoided the problem by using network science metrics to show hierarchy structure and modularity in protein interactomes (Ravasz et al., 2002; Barabási and Oltvai, 2004). Partitioning the interactome into workable subnetworks is an alternative strategy that has demonstrated the presence of other topological motifs, such as protein complexes and hubs. Still, network science metrics provide only a high-level topological understanding, while partitioning the interactome into subnetworks will never provide a complete overview in detail.

In this study, we characterized the topology of a physical PPI interactome from a conventional dendritic cell lineage 1 (cDC1), consisting of 8,569 proteins and 278,365 PPIs. Specifically, we visualized the interactome's complete topology in a weighted topological overlap (wTO) affinity matrix and decoded the substructures observed in this matrix into topological network motifs. In the cDC1 PPI interactome, we found rare topological motifs such as the bow-tie motif, which connected protein complexes, and a type of "multi-protein hub" motif centered around Rho-family GTPases, while we also observed well-known motifs such as protein complexes and hub proteins. Extensive network and functional analysis of the cDC1 bow-tie motifs' key proteins, called "knot" proteins (n = 294), revealed that they were a pervasive topological trend, having multifunctional capabilities, specific network properties, and widespread expression in cells and tissue, and were important for cell survival.

RESULTS

Construction and Visualization of the cDC1 Interactome

We sorted XCR1^{+}MHCII^{hi} cDC1 cells using fluorescence-activated cell sorting (FACS) extracted from murine mesenteric lymph nodes (Figure S1). Then we generated gene expression RNA sequencing (RNA-seq) data and determined which genes were actively expressed in cDC1 cells (see STAR Methods for details). For the creation of the cDC1 PPI interactome, we obtained PPIs from the meta-database InWeb_IM, which contains experimentally generated, scored, and benchmarked PPIs from human, mice, and other eukaryotes (Li et al., 2017). The cDC1-specific interactome were constructed using the node-removal approach applied to InWeb_IM, including the protein products of expressed genes in cDC1 cells and their associated PPIs (Magger et al., 2012). This resulted in an interactome consisting of 8,569 proteins and 278,365 PPIs with a mean shortest path of 2.68 interactions.

To decode the topology of the cDC1-specific interactome, we calculated the pairwise topological relationship between all 8,569 proteins (~36.7 million pairs) using the wTO measure. The wTO is the relative number of shared first-order neighbors between two proteins and is a good measure of ''interconnectedness'' (Ravasz et al., 2002; Zhang and Horvath, 2005) (Figure 1B). Hence, the wTO measure between two proteins captures more network topology information than the confidence score of their PPI interaction. When calculating the wTO measure, we weighed each PPI according to its confidence score (0.0–1.0), which is a score assigned by the InWeb_IM database (see STAR Methods). We visualized the wTO pairs in a hierarchically clustered wTO affinity matrix, which we combined with a half black-and-white matrix of direct PPIs (Figure 1A).

An Approach to Decode Network Topology

The cDC1-specific wTO affinity matrix contains substructures in the form of dark lines, squares, dots, and more (Figure 1A). This substructure encodes the network topology; hence, the topology can be decoded directly from the matrix. To demonstrate, we generated a mock network and its corresponding direct PPI/wTO matrix (Figures 1C and 1D). The mock network illustrates how two types of topological motifs (i.e., hubs and complexes) create two distinct types of substructures in the wTO matrix (lines and squares). Thus, we can determine that the mock network contains five complexes, all connected by a hub, solely by reading the matrix. For very large networks, this decoding approach is especially useful, as such networks cannot be visualized in a comprehensible manner using nodes and edges (i.e., graph visualization). We provide an R package for the construction of wTO matrices directly from igraph objects, applicable to any weighted or unweighted network (https://github.com/k-niss/bowtie).

Topological Characterization of the cDC1 Interactome

Manual inspection of the cDC1 interactome's direct PPI/wTO matrix (Figure 1A) revealed four types of substructures (examples in Figures 2B–2I). Similar to the mock network, we found square substructures that encoded almost or fully connected protein complexes (Figures 2B and 2C), while line substructures encoded hub proteins, here exemplified by SUMO3 (Figures 2D and 2E). We further noticed that there were line substructures between the squares in multiple cases (Figure 2F). When such areas of the matrix were graph-visualized (Figure 2G), the substructures turned out to be a topological motif called a bow-tie, which connected the protein complexes (Broder et al., 2000; Csete and Doyle, 2004; Kitano, 2004, 2007; Tieri et al., 2010). In its simplest form, a bow-tie motif consists of a ''knot'' in the center, connected to two separate entities. In our case, the ''knot'' is a protein connected to two protein complexes via interaction fans (Figure 2G, schematic). Finally, we observed a block substructure (Figure 2H), which encoded a ''multi-protein hub'' motif (Figure 2I). In this motif, a core set of 14 Rho-family GTPases (Figure 2I, blue nodes) all acted as hub proteins for the same set of 70 proteins (Figure 2I, gray nodes). The set of 70 proteins were only sparsely interconnected themselves (see detailed graph visualization in Figure S1). It was clear that the square substructures (complexes), line substructures (hubs), and lines between square substructures (bow-ties) were reoccurring throughout the cDC1 matrix (Figure 1A). However, the ''multi-hub protein'' motif was observed only once.

Protein complexes and hub proteins have been extensively studied at an interactome-wide level, using community detection algorithms to find protein complexes and using the high degree as a selection criterion to find hub proteins. In contrast, bow-tie motifs have been detected only a few times in PPI networks (Oda and Kitano, 2006; Polouliakh et al., 2009; Abd-Rabbo and Michnick, 2017), and their potential role as recurring connectors of protein complexes has not been investigated. Therefore, to obtain an overview of the bow-tie motifs in the cDC1 interactome, we cataloged the ''knot'' proteins and their interaction fans using a two-step process. First, we computationally traversed the cDC1 interactome's direct PPI matrix along the diagonal for square substructures and identified 64 protein complexes ranging from 9 to 132 proteins in size (Figure S2; STAR Methods). All protein complexes were densely interconnected (mean clustering coefficient 0.9549 ± 0.06). We annotated each protein complex with its most significant term from either the Gene Ontology (GO) Biological Process Database or the Reactome Database (Table S1A). Second, we computationally searched the cDC1 interactome's direct PPI matrix for line substructures between the identified protein complexes (Figure S2; STAR Methods). The ''lines'' represent interactions between a ''knot'' protein and the members of a protein complex. Furthermore, to be classified as a ''knot'' protein, a protein had to be connected to $\geq$2 protein complexes with at least 10 PPIs connecting it to each complex, where the average confidence score of the PPIs had to be $\geq$0.9 (see STAR Methods). Here, confidence score refers to the metric assigned to each PPI by the InWeb_IM database. Thereby, we identified 294 ''knot'' proteins (Table S1B).

Network Analysis of the Bow-Tie Motifs

We performed a network analysis of the bow-tie motifs. One third of the ''knot'' proteins had more than two interaction fans (Figure 3A). ''Knot'' proteins also had more PPIs than the typical protein in the cDC1-specific interactome (Figure 3B). By determining network metrics that describe how important a protein is for the global information flow of the interactome, we found

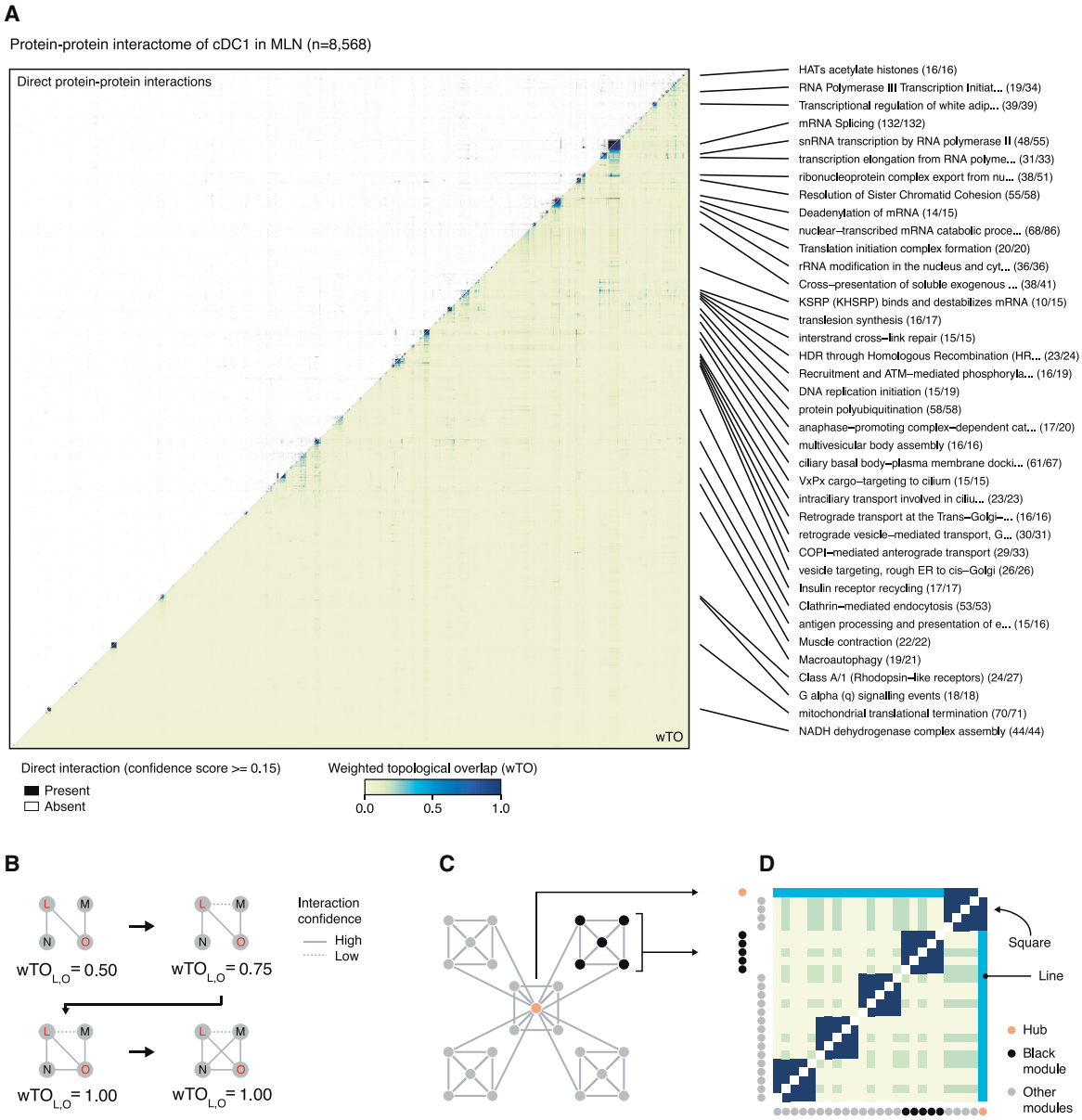


Figure 1. Network Topology of the cDC1-Specific Interactome

(A) In the upper/left matrix, direct protein-protein interactions from InWeb_IM are displayed in black. In the lower/right matrix, the weighted topological overlap (wTO) measure between protein pairs is displayed. Larger protein complexes ($n \geq 15$) are annotated to the right using the Gene Ontology Biological Process database or Reactome. The numbers in parentheses are the proteins significant for the annotation term relative to the complex size. The matrix is hierarchically clustered on the basis of the wTO measure.

(B) Graphical representation of the wTO measure. The wTO measure increases as proteins become more interconnected, here exemplified by protein L and O: $wTO_{L,O}$ increases to 0.75 when L-M connects and further increases to 1.0 when also N-O connects. Yet $wTO_{L,O}$ does not increase further by adding the N-M connection, because this connection does not increase the inter-connection between L and O.

(C) A mock network consisting of two types of topological motifs: a hub (orange node) and five complexes (gray/black nodes).

(D) The direct PPI matrix combined with the wTO matrix of the mock network. The orange hub motif is observable as long horizontal and vertical lines in the wTO matrix, while the complexes appear as dark blue squares. The network topology can be read directly from the matrix by linking specific matrix substructures to known topological motifs.

See also Figure S1 and Table S1A.

that "knot" proteins were less important than other proteins with a similarly large number of PPIs (i.e., degree distribution) (see STAR Methods), as their metrics were significantly lower on average: betweenness centrality p = 0.0161, closeness centrality p = 0.0001, and nodal efficiency p = 0.0e−4 (Figure 3C). In contrast, we found "knot" proteins significantly more tightly

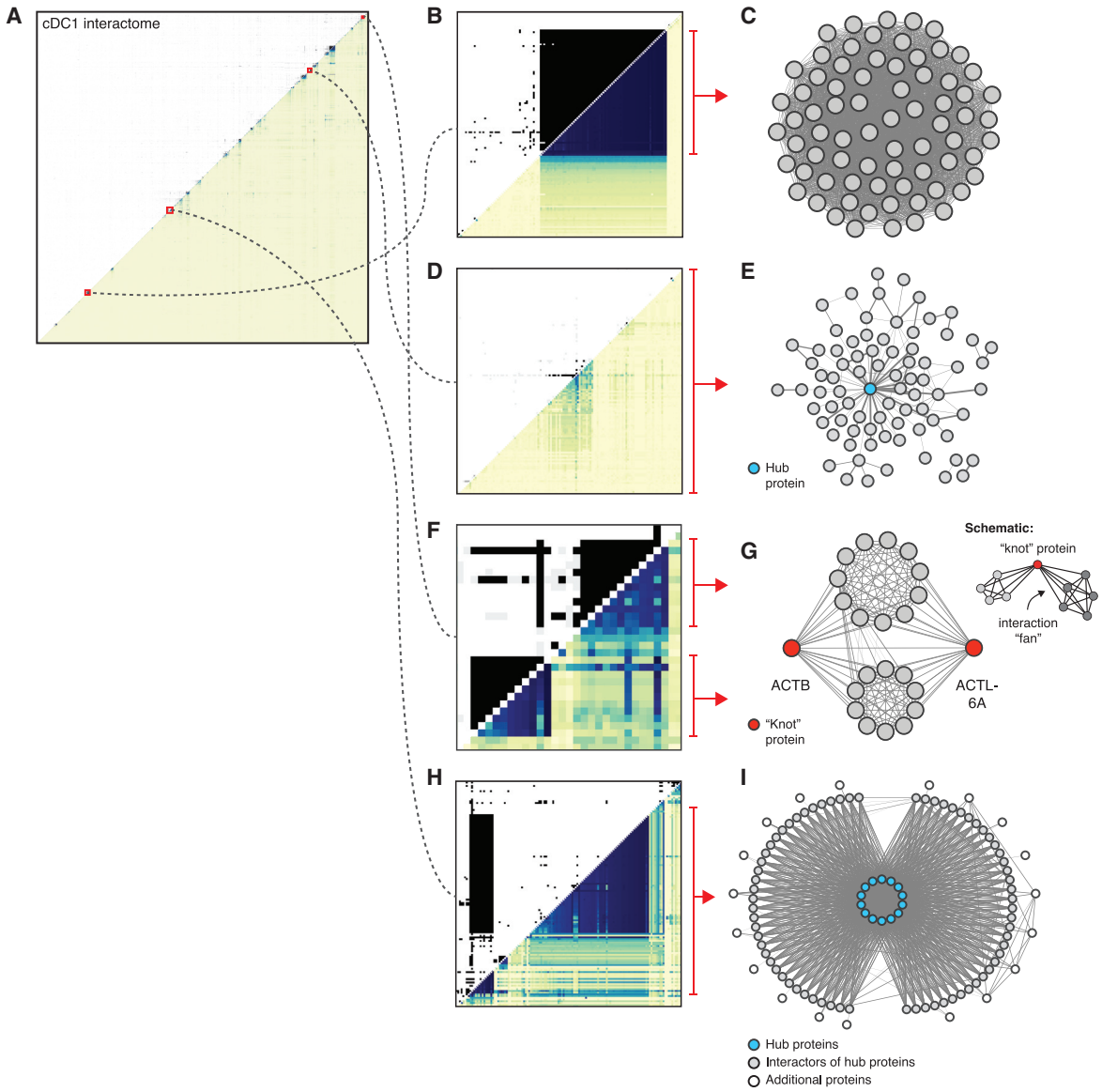


Figure 2. Overview of Different Matrix Substructures and Corresponding Topology Motifs Found in the cDC1 Interactome PPI/wTO Matrix

(A) The location of the discovered substructure examples in the complete cDC1 PPI/wTO matrix.

(B, D, F, and H) The substructures were (B) squares, (D) horizontal and vertical lines, (F) lines between squares, and (H) lines alongside each other forming a block.

(C, E, G, and I) By extracting the proteins from the respective matrices in (B), (D), (F), and (H), as indicated by the red arrows, and plotting their interactions, (C) the squares were found to encode protein complexes, (E) the lines encoded hub proteins, (G) the lines between squares encoded bow-tie motifs connecting protein complexes, and (I) the blocks encoded a type of "multi-protein hub" motif.

See also Figure S1.

clustered in their local network neighborhood compared with other proteins with a similar degree distribution, as described by the high clustering coefficient ($p = 0.0e{-}4$) (Figure 3C).

We constructed a connection grid to illustrate how the "knot" proteins connected the 64 protein complexes (Figure 3D). The grid showed that bow-tie motifs collectively create a scaffold that connects most of the protein complexes in the cDC1 interactome (51 of 64). We colored the protein complexes by their subcellular compartment (Table S1C), which revealed that complexes with related functions had more connections than the reverse. For an example, two protein complexes of the Golgi apparatus had 50 bow-tie motifs connecting them (Figure 3D, boxed). Interestingly, we further observed examples of bow-tie motifs that cross subcellular compartments, such as the motifs created by "knot" protein HSPA8 or ATP6AP1 (Figure 3D, asterisks).

Functional Analysis of the Bow-Tie Motif

We became interested in the functional rationale behind connecting a majority of protein complexes using bow-tie motifs.

CellPress
OPEN ACCESS

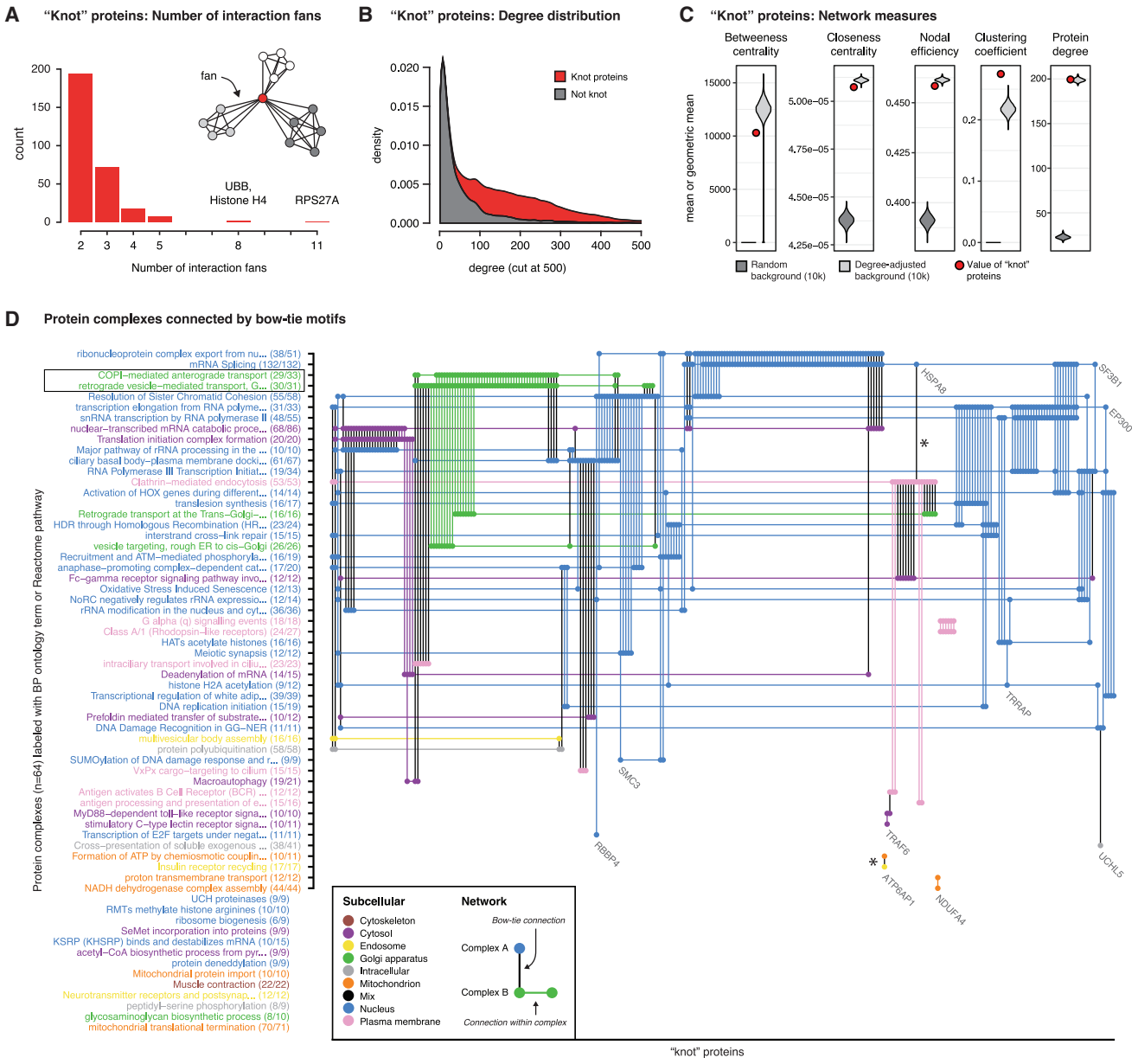


Figure 3. The "Knot" Proteins of Bow-Tie Motifs

(A) The number of interactions fans per "knot" protein.

(B) Degree distribution for "knot" proteins (red) and non-knot proteins (gray).

(C) Mean network measures for the "knot" proteins (red) against a background of proteins sampled at random (dark gray) and sampled at random while keeping a similar degree distribution (light gray).

(D) A grid of protein complexes connected by bow-tie motifs. The 294 identified "knot" proteins are hierarchically ordered along the x axis, and annotated protein complexes are similarly ordered along the y axis. Vertical lines with dots show which protein complexes are connected by "knot" proteins. Horizontal lines help the eye in tracing the protein complexes across the plot. The lengths of lines have no meaning. A "knot" protein can connect more than two protein complexes; hence more than two dots can occur on a vertical line. Indented protein complexes (n = 13) on the y axis (bottom part) are not connected by bow-ties.

See also Figure S2 and Tables S1A–S1C.

We investigated the biological functions of the 294 "knot" proteins and discovered that they were associated with a significantly large number of GO biological processes relative to other proteins with a similar degree distribution (permutation test, 10,000 degree-adjusted sampling) (Figure 4A). This result suggests that the "knot" proteins are multifunctional, which is perhaps enabled by their bow-tie topology.

Investigating the bow-tie motifs in detail, we found multiple examples of multifunctional "knot" proteins, whose biological function appears to change according to which protein complex

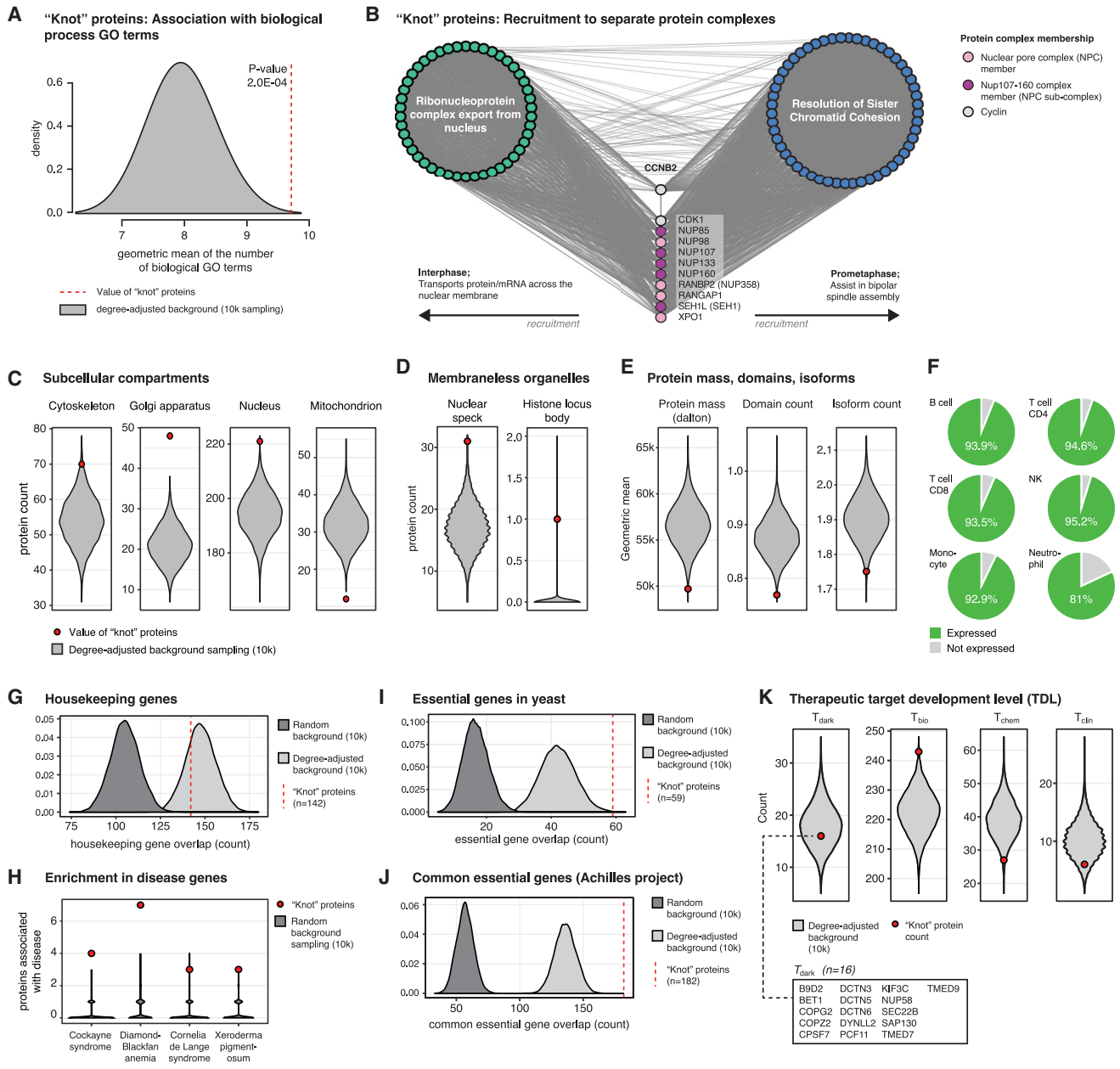


Figure 4. Functional Characterization of Bow-Ties and "Knot" Proteins

(A) The geometric mean of the number of associated biological process Gene Ontology (GO) terms for each "knot" protein (red line) relative to a distribution of 10,000 random protein sets with a degree distribution similar to the "knot" proteins (gray density).

(B) A network visualization of the protein complexes "ribonucleoprotein complex export from nucleus" and "resolution of sister chromatid cohesion" with 11 "knot" proteins connecting them via bow-tie motifs.

(C and D) The number of "knot" proteins in different subcellular compartments (C) and in membraneless organelles (D) relative to a distribution of 10,000 random, degree-adjusted protein sets.

(E) The geometric mean of the protein mass, domain count and isoform count for each "knot" protein relative to 10,000 random, degree-adjusted protein sets.

(F) The percentage of the "knot" proteins found to be expressed in six different immune cells.

(G) The number of "knot" proteins found in a set of housekeeping genes, relative to a distribution of 10,000 randomly sampled protein sets (dark gray) or sampled using a degree-adjusted approach (light gray).

(H) The number of "knot" protein in disease protein sets related to Cockayne syndrome, Diamond-Blackfan anemia, Cornelia de Lange syndrome, and xeroderma pigmentosum relative to a distribution of 10,000 randomly sampled protein sets.

(legend continued on next page)

they interact with. One such example was the 11 "knot" proteins connecting the protein complexes "ribonucleoprotein complex export from nucleus" and "resolution of sister chromatid cohesion" (Figure 4B). Four of the "knot" proteins are members of the nuclear pore complex (NPC), 5 of them members of the sub-nucleoporin complex Nup107–160 (Harel et al., 2003), and 2 of them are cyclins. During cellular interphase, the NPC and Nup107–160 complexes form pores in the nuclear membrane through which RNA and proteins are transported (Griffis et al., 2002), and they are therefore part of the "ribonucleoprotein complex export from nucleus" complex. In contrast, during cellular prometaphase, the NPC and Nup107–160 complexes participate in bipolar spindle assembly (Orjalo et al., 2006), which is a process associated with the "resolution of sister chromatid cohesion" complex. Importantly, the biological processes of "ribonucleoprotein complex export from nucleus" and "resolution of sister chromatid cohesion" are temporally non-overlapping, because the nuclear membrane is intact in interphase, while it is dissolved during prometaphase. The example illustrates how the "knot" proteins are potentially multifunctional, carrying out different functions in collaboration with different protein complexes dependent on the cellular context. We present two additional example of multifunctional "knot" proteins in Figure S3 related to the proteasome, the INO80 helicase complex and two types of DNA repair (Gillet and Schärer, 2006; Salas et al., 2009; Svendsen et al., 2009; Yao et al., 2008; Zediak and Berger, 2008).

Function, Expression, and Clinical Significance of "Knot" Proteins

We continued by performing a broader functional characterization of the 294 "knot" proteins. When assigning corresponding subcellular compartments to each "knot" protein using the COMPARTMENTS database (Binder et al., 2014), we found that "knot" proteins were significantly enriched in the cytoskeleton ($p = 0.012$), the Golgi apparatus ($p = 0.0e{-}4$), and the nucleus ($p = 1.0e{-}4$), while being underrepresented in the mitochondria ($p = 0.0e{-}4$) (permutation test, 10,000 degree-adjusted sampling) (Figure 4C). Membraneless organelles have recently gained interest as another strategy for intracellular compartmentalization complementing membrane-enclosed organelles (Nott et al., 2016). To investigate the importance of bow-tie topology in these, we calculated the enrichment of "knot" proteins in ten membraneless organelles (Gomes and Shorter, 2019). "Knot" proteins were found to be significantly enriched in two membraneless organelles: nuclear specks ($p = 4e{-}04$) and in histone locus bodies ($p = 0.0406$) (permutation test, 10,000 degree-adjusted sampling) (Figure 4D). We further checked for enrichment of membraneless organelle proteins in our 64 protein complexes, as the complexes resemble membraneless organelles on a number of parameters (Gomes and Shorter, 2019). Protein sets from five of ten membraneless organelles were found to be enriched in a protein complex, whereof three of the organelles were enriched in more than two protein complexes (Figure S3).

Investigating the physical properties of "knot" proteins, we found that they had a significantly lower protein mass ($p = 1.0e{-}3$), fewer domains ($p = 6e{-}4$), and fewer isoforms ($p = 6.3e{-}3$) compared with other high-degree proteins. However, the sizes of the differences were small in all three cases (Figure 4E).

Although we worked on an interactome specific to cDC1, our analyses indicated that the "knot" proteins themselves were not cDC1 specific. We thus investigated the expression of the "knot" proteins in six types of human immune cells using a separate dataset (Linsley et al., 2014) and found that the majority (81%–95.2%) of the "knot" proteins were expressed (Figure 4F). We furthermore created a B cell interactome (n = 9,919 proteins) and a natural killer (NK) cell interactome (n = 10,048 proteins) on the basis of the mentioned dataset (Linsley et al., 2014) and undertook a complete topological analysis on both of them identical to the one applied to the cDC1 interactome (Figure S4). In the B cell, we found 73 protein complexes and 392 "knot" proteins, while in the NK cell, we found 68 protein complexes and 361 "knot" proteins (Tables S1D–S1G). When comparing the cells, 223 "knot" proteins were shared among B, NK, and cDC1 cells, suggesting that many "knot" proteins are not cell type specific. Moreover, using the tissue database GTEx (Battle et al., 2017), we found 70% of the cDC1 "knot" proteins expressed at five transcripts per million in all GTEx tissues (Figure S3). Collectively, these results indicate that a major part of the cDC1 "knot" proteins are common proteins, being expressed in many different cell types and tissues. Testing for the "knot" proteins' enrichment in a set of housekeeping genes (Eisenberg and Levanon, 2013) (Figure 4G), the "knot" proteins were found to be enriched relative to a random sampling of proteins ($p = 0.0e{-}04$), but when compared with other high-degree proteins, the number was expected.

With the unique network features of "knot" proteins and their expression across cell types and tissues, we investigated their clinical significance. Using the DISEASE database (Pletscher-Frankild et al., 2015), we found "knot" proteins significantly enriched (permutation test, 10,000 sampling) in four severe disorders (Figure 4H). These were related to mutations in either DNA damage repair genes (Cockayne syndrome, Cornelia de Lange syndrome, xeroderma pigmentosum) or ribosomal genes (Diamond-Blackfan anemia). We found it striking that no cancer types were found, as cancer often increases the mutation rate by mutating DNA repair genes. A potential reason for the lack of cancer enrichment could be that many "knot" proteins are fundamental for cell survival (i.e., essential). We tested this by asking if the "knot" proteins were enriched (permutation test, 10,000 sampling) in a set of *S. cerevisiae* essential genes (Liu et al., 2015) (Figure 4I) and in a set of common essential genes

(I and J) The number of "knot" proteins found in a set of essential genes from *S. cerevisiae* (I) and in a set of common essential genes from Project Achilles (J), relative to a distribution of 10,000 randomly sampled protein sets (dark gray) or sampled using a degree-adjusted approach (light gray).
(K) The "knot" protein count in each of the four therapeutic target development levels (see text for details) relative to 10,000 degree-adjusted sampled protein sets. Proteins found in T_{dark}, which are the least studied, have been highlighted in a box.
See also Figures S3 and S4 and Tables S1D–S1G.

from the Achilles project (Figure 4J), which is a screen of 18,333 knocked-out genes across 625 human cell lines. The "knot" proteins were found to be enriched in both the *S. cerevisiae* set (random background, p = 0.0e−4; degree-adjusted background, p = 0.0e−4) and in the Achilles project set (random background, p = 0.0e−4; degree-adjusted background, p = 0.0e−4).

Finally, we looked up the therapeutic target development level (TDL) of each "knot" protein (Oprea et al., 2018) (Figure 4K). This is a knowledge metric that assess how well studied a protein is. The TDL levels are ranked from least to most studied: T_{dark}, T_{bio}, T_{chem}, and T_{clin} (See Oprea et al., 2018, for the exact definition of each T level). The "knot" proteins were enriched in T_{bio} (p = 0.0015) and underrepresented in T_{chem} (p = 0.0125). Although the "knot" protein were not enriched in T_{dark}, we would still like to highlight the 16 "knot" proteins found in this category, as they represent a particularly interesting group for further studies.

DISCUSSION

Through a complete visual overview of a cDC1-specific PPI interactome's network topology, we discovered that bow-tie motifs connect protein complexes in cellular interactomes. A set of 294 bow-tie motifs connected 51 densely interconnected protein complexes in the cDC1 interactome, while similar trends were shown within a B cell and an NK cell interactome. The bow-tie motif is known for being the design of the World Wide Web (Broder et al., 2000) and is used in systems engineering (Tieri et al., 2010). In cell biology, bow-tie motifs have been shown to exist in the Toll-like receptor pathway, G protein-coupled receptor pathway, and kinase/phosphatase signaling pathways (Abd-Rabbo and Michnick, 2017; Oda and Kitano, 2006; Polouliakh et al., 2009). With this work, we demonstrate that bow-tie motifs are present in diverse functional areas of the cellular interactome, not just in signaling pathways, and we also show that the bow-tie motif is far more numerous than previously thought.

Clearly, the bow-tie motif is a deviation of the more general hub protein motif, as it is centered around a protein with many PPIs. As such, the bow-tie motif could be considered a subtype of the hub protein, as it has many PPIs, but it also has a strict motif definition: it must have a large number of PPIs connecting it to at least two protein complexes. The bow-tie motif furthermore grants its "knot" proteins statistical significantly different network properties relative to other high-degree proteins (i.e., hub proteins), being less important for global network information flow (low betweenness centrality, low closeness centrality, and low nodal efficiency) but highly clustered into its local interactome neighborhood (high clustering coefficient). Within this local interactome neighborhood, the "knot" protein is then, conversely, very important for information flow. We find multiple examples of bow-tie motifs with at least three interaction fans, which has not been seen in a PPI interactome before. We also find that some protein complex pairs are connected by up to 50 bow-tie motifs in a repeated fashion, which has not been associated with bow-tie motifs before either.

We showed multiple examples of bow-tie motifs whose "knot" proteins carried out certain biological processes when associated with one complex, while carrying out functionally distinct biological processes when associated with another complex. A statistical analysis of all "knot" proteins further revealed that in general they were associated with a significantly larger number of biological processes than other high-degree proteins. In sum, our results suggest that "knot" proteins tend to be multifunctional. The network motif of the bow-tie possibly enables the "knot" proteins to carry out multiple functions by allowing them to switch between protein complexes. This multifunctionality role has not been linked to bow-tie motifs before, as they have previously been found only in signaling pathways, where they are known to coordinate signal (Kitano, 2004). However, investigation into "knot" proteins that are not currently known to be multifunctional is required before the "knot" protein designation can be used as an indicator of multifunctionality.

We found that "knot" proteins were enriched in the nucleus, Golgi apparatus, and cytoskeleton. In the nucleus, many distinct biological processes apply the same mechanistic steps, such as DNA helix unwinding, which is used to initiate DNA replication, DNA repair, and transcription. We hypothesized that "knot" proteins were situated in bow-tie motifs, such that their function could be shared among distinct biological processes. It hereby follows that "knot" proteins are overrepresented in the nucleus because sharing of mechanistic steps and therefore enzymatic proteins is common in the nucleus. Similarly, in the Golgi apparatus, the vesicle transport and trafficking system consists of multiple biological processes, including antero- and retrograde transport between the endoplasmic reticulum (ER) and Golgi and vesicle transport to endosomes, lysosomes, and extracellular space. As in the nucleus, these processes share proteins for common mechanistic steps, and enrichment of multifunctional "knot" proteins is therefore not surprising. In contrast, we showed underrepresentation of "knot" proteins in the mitochondrion. This may be explained by evolutionary theory, wherein the mitochondrion is thought to be a descendent of an endocytosed bacterium. The protein machinery of the mitochondrion is therefore expected to be somewhat isolated from the remaining cell.

The "knot" proteins were found to be overrepresented in protein sets of four severe diseases that are caused by mutations in DNA repair pathways. Given the severity of these diseases, we investigated and found that "knot" proteins were enriched in two sets of genes known to be essential. The multifunctional feature of "knot" proteins may be the driving reason for their essentiality, as multifunctionality is a core principle of cellular organization. This is demonstrated by the human genome's surprisingly small number of protein-coding genes. The number is in fact still decreasing in reference genome revisions today, as some predicted genes and intron-free open reading frames are found to be non-coding (Ezkurdia et al., 2014).

We observed a "multi-hub protein" motif of 14 Rho-family GTPases acting as hubs for the same set of 70 proteins. The Rho family of GTPases regulates intracellular actin dynamics, such as the formation of lamellipodium on cells (Hall, 2012). Other studies have shown that RhoA, RhoB, RhoC, RhoD, and Cdc42 of the Rho family have a hub-like motif (Paul et al., 2017; Pertz et al., 2008). Our network expands on this, illustrating that more Rho-family GTPases have hub-like motif topology in the PPI interactome than previously thought and that they share 70 protein interactors in the case of the cDC1 interactome.

We apply hierarchical clustering and the proven wTO measure of protein pairs to a new problem: the analysis of network topology in large cellular interactomes. Our work further presents an approach to decode substructures of a wTO matrix into topological motifs, thereby revealing fundamental topological trends. A key strength of this methodology is that by visualizing the cellular interactomes' complete network topology up front, we were able to start our topological characterization with no prior assumptions regarding the network's structure. This feature sets our approach apart from other network analysis algorithms. Furthermore, we chose to study the cDC1, B cell, and NK cell because of the availability of data, while a study of the unspecific PPI interactome was not included, as its biological relevance is limited.

Cell type-specific protein interactomes are incredibly complex, often containing more than 10,000 proteins that form dense interaction networks, typically with an average shortest path of two or three steps. This complexity makes the study of their topology difficult. Our approach to decode their topology and uncover topological motifs and trends will hopefully increase the use of interactomes and networks in general. Furthermore, our discovery and characterization of the bow-tie motif in the context of the cellular PPI interactome provides fundamental knowledge on how the intracellular system is organized and operates.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials Availability
  - Data and Code Availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Cell preparation
  - Flow cytometry analysis and cell sorting
  - Library preparation
  - mRNA sequencing using Illumina
  - RNA-seq data analysis of cDC1 data
  - Analysis of sorted immune cell RNA-seq data
  - The cDC1-specific protein interactome
  - The cDC1-specific interactome topology
  - Matrix and network visualization
  - Bow-tie motif localization in wTO matrix
  - Protein complex localization in wTO matrix
  - DISEASES and COMPARTMENTS databases
  - Membraneless organelles
  - Protein mass, domains and isoforms
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.celrep.2020.107763.

ACKNOWLEDGMENTS

RNA-seq sample libraries were prepared by Julien Vandamme, PhD, Mucosal Immunology, Technical University of Denmark. RNA-seq was performed at the Genomics Core Facility, KFB: Center of Excellence for Fluorescent Bioanalytics, University of Regensburg, Regensburg, Germany. Associate Professor Jose MG Izarzugaza, PhD, DTU Bioinformatics, Technical University of Denmark, provided expertise for the degree-adjusted permutation tests. This work was supported by the Lundbeck Foundation, Denmark (grant agreement 110145, RIMMI) and the Novo Nordisk Foundation, Denmark (grant agreement NNF14CC0001 and grant agreement NNF17OC0027594, Challenge). C.G.-C. was supported by a postdoctoral fellowship from the Carl Tryggers Foundation, Sweden.

AUTHOR CONTRIBUTIONS

Conceptualization, Methodology, Software, Formal Analysis, Writing, and Visualization, K.N.; Investigation and Resources, C.G.-C. and T.J.; Conceptualization and Writing – Review & Editing, J.X.H.; Resources, W.W.A.; Writing – Review & Editing, K.G.B.; Conceptualization, Writing – Review & Editing, Supervision, and Funding Acquisition, S.B.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: August 16, 2019
Revised: February 3, 2020
Accepted: May 21, 2020
Published: June 16, 2020

REFERENCES

Abd-Rabbo, D., and Michnick, S.W. (2017). Delineating functional principles of the bow tie structure of a kinase-phosphatase network in the budding yeast. BMC Syst. Biol. *11*, 38.

Andrews, S. (2010). FastQC. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Barabási, A.-L., and Oltvai, Z.N. (2004). Network biology: understanding the cell's functional organization. Nat. Rev. Genet. *5*, 101–113.

Battle, A., Brown, C.D., Engelhardt, B.E., and Montgomery, S.B.GTEx Consortium; Laboratory, Data Analysis &Coordinating Center (LDACC)—Analysis Working Group; Statistical Methods Groups—Analysis Working Group; Enhancing GTEx (eGTEx) Groups; NIH Common Fund; NIH/NCI; NIH/NHGRI; NIH/NIMH; NIH/NIDA; Biospecimen Collection Source Site—NDRI; Biospecimen Collection Source Site—RPCI; Biospecimen Core Resource—VARI; Brain Bank Repository—University of Miami Brain Endowment Bank; Leidos Biomedical—Project Management; ELSI Study; Genome Browser Data Integration &Visualization—EBI; Genome Browser Data Integration &Visualization—UCSC Genomics Institute, University of California Santa Cruz; Lead Analysts; Laboratory, Data Analysis &Coordinating Center (LDACC); NIH Program Management; Biospecimen Collection; Pathology; eQTL Manuscript Working Group (2017). Genetic effects on gene expression across human tissues. Nature *550*, 204–213.

Binder, J.X., Pletscher-Frankild, S., Tsafou, K., Stolte, C., O'Donoghue, S.I., Schneider, R., and Jensen, L.J. (2014). COMPARTMENTS: unification and visualization of protein subcellular localization evidence. Database (Oxford) *2014*, bau012.

Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., Tomkins, A., and Wiener, J. (2000). Graph structure in the Web. Comput. Netw. *33*, 309–320.

Chatr-Aryamontri, A., Oughtred, R., Boucher, L., Rust, J., Chang, C., Kolas, N.K., O'Donnell, L., Oster, S., Theesfeld, C., Sellam, A., et al. (2017). The

BioGRID interaction database: 2017 update. Nucleic Acids Res. *45* (*D1*), D369–D379.

Chua, H.N., Sung, W.-K., and Wong, L. (2006). Exploiting indirect neighbours and topological weight to predict protein function from protein-protein interactions. Bioinformatics *22*, 1623–1630.

Csete, M., and Doyle, J. (2004). Bow ties, metabolism and disease. Trends Biotechnol. *22*, 446–450.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Eisenberg, E., and Levanon, E.Y. (2013). Human housekeeping genes, revisited. Trends Genet. *29*, 569–574.

Ezkurdia, I., Juan, D., Rodriguez, J.M., Frankish, A., Diekhans, M., Harrow, J., Vazquez, J., Valencia, A., and Tress, M.L. (2014). Multiple evidence strands suggest that there may be as few as 19,000 human protein-coding genes. Hum. Mol. Genet. *23*, 5866–5878.

Gillet, L.C.J., and Schärer, O.D. (2006). Molecular mechanisms of mammalian global genome nucleotide excision repair. Chem. Rev. *106*, 253–276.

Gomes, E., and Shorter, J. (2019). The molecular language of membraneless organelles. J. Biol. Chem. *294*, 7115–7127.

Griffis, E.R., Altan, N., Lippincott-Schwartz, J., and Powers, M.A. (2002). Nup98 is a mobile nucleoporin with transcription-dependent dynamics. Mol. Biol. Cell *13*, 1282–1297.

Hall, A. (2012). Rho family GTPases. Biochem. Soc. Trans. *40*, 1378–1382.

Harel, A., Orjalo, A.V., Vincent, T., Lachish-Zalait, A., Vasu, S., Shah, S., Zimmerman, E., Elbaum, M., and Forbes, D.J. (2003). Removal of a single pore subcomplex results in vertebrate nuclei devoid of nuclear pores. Mol. Cell *11*, 853–864.

Hishigaki, H., Nakai, K., Ono, T., Tanigami, A., and Takagi, T. (2001). Assessment of prediction accuracy of protein function from protein–protein interaction data. Yeast *18*, 523–531.

Hu, J.X., Thomas, C.E., and Brunak, S. (2016). Network biology concepts in complex disease comorbidities. Nat. Rev. Genet. *17*, 615–629.

Kitano, H. (2004). Biological robustness. Nat. Rev. Genet. *5*, 826–837.

Kitano, H. (2007). A robustness-based approach to systems-oriented drug design. Nat. Rev. Drug Discov. *6*, 202–210.

Langfelder, P. (2012). flashClust: implementation of optimal hierarchical clustering. https://rdrr.io/cran/flashClust/.

Law, C.W., Chen, Y., Shi, W., and Smyth, G.K. (2014). voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. *15*, R29.

Lefebvre, C., Rajbhandari, P., Alvarez, M.J., Bandaru, P., Lim, W.K., Sato, M., Wang, K., Sumazin, P., Kustagi, M., Bisikirska, B.C., et al. (2010). A human B-cell interactome identifies MYB and FOXM1 as master regulators of proliferation in germinal centers. Mol. Syst. Biol. *6*, 377.

Li, T., Wernersson, R., Hansen, R.B., Horn, H., Mercer, J., Slodkowicz, G., Workman, C.T., Rigina, O., Rapacki, K., Stærfeldt, H.-H., et al. (2017). A scored human protein-protein interaction network to catalyze genomic interpretation. Nat. Methods *14*, 61–64.

Licata, L., Briganti, L., Peluso, D., Perfetto, L., Iannuccelli, M., Galeota, E., Sacco, F., Palma, A., Nardozza, A.P., Santonico, E., et al. (2012). MINT, the molecular interaction database: 2012 update. Nucleic Acids Res. *40*, D857–D861.

Linsley, P.S., Speake, C., Whalen, E., and Chaussabel, D. (2014). Copy number loss of the interferon gene cluster in melanomas is linked to reduced T cell infiltrate and poor patient prognosis. PLoS ONE *9*, e109760.

Liu, G., Yong, M.Y.J., Yurieva, M., Srinivasan, K.G., Liu, J., Lim, J.S.Y., Poidinger, M., Wright, G.D., Zolezzi, F., Choi, H., et al. (2015). Gene essentiality is a quantitative property linked to cellular evolvability. Cell *163*, 1388–1399.

Luda, K.M., Joeris, T., Persson, E.K., Rivollier, A., Demiri, M., Sitnik, K.M., Pool, L., Holm, J.B., Melo-Gonzalez, F., Richter, L., et al. (2016). IRF8 transcription-factor-dependent classical dendritic cells are essential for intestinal T cell homeostasis. Immunity *44*, 860–874.

Magger, O., Waldman, Y.Y., Ruppin, E., and Sharan, R. (2012). Enhancing the prioritization of disease-causing genes through tissue specific protein interaction networks. PLoS Comput. Biol. *8*, e1002690.

Nott, T.J., Craggs, T.D., and Baldwin, A.J. (2016). Membraneless organelles can melt nucleic acid duplexes and act as biomolecular filters. Nat. Chem. *8*, 569–575.

Oda, K., and Kitano, H. (2006). A comprehensive map of the toll-like receptor signaling network. Mol. Syst. Biol. *2*, 2006.0015.

Oprea, T.I., Bologa, C.G., Brunak, S., Campbell, A., Gan, G.N., Gaulton, A., Gomez, S.M., Guha, R., Hersey, A., Holmes, J., et al. (2018). Unexplored therapeutic opportunities in the human genome. Nat. Rev. Drug Discov. *17*, 317–332.

Orjalo, A.V., Arnaoutov, A., Shen, Z., Boyarchuk, Y., Zeitlin, S.G., Fontoura, B., Briggs, S., Dasso, M., and Forbes, D.J. (2006). The Nup107-160 nucleoporin complex is required for correct bipolar spindle assembly. Mol. Biol. Cell *17*, 3806–3818.

Paul, F., Zauber, H., von Berg, L., Rocks, O., Daumke, O., and Selbach, M. (2017). Quantitative GTPase affinity purification identifies Rho family protein interaction partners. Mol. Cell. Proteomics *16*, 73–85.

Pedersen, H.K., Gudmundsdottir, V., and Brunak, S. (2017). Pancreatic islet protein complexes and their dysregulation in type 2 diabetes. Front. Genet. *8*, 43.

Pertz, O.C., Wang, Y., Yang, F., Wang, W., Gay, L.J., Gristenko, M.A., Clauss, T.R., Anderson, D.J., Liu, T., Auberry, K.J., et al. (2008). Spatial mapping of the neurite and soma proteomes reveals a functional Cdc42/Rac regulatory network. Proc. Natl. Acad. Sci. U S A *105*, 1931–1936.

Pletscher-Frankild, S., Pallejà, A., Tsafou, K., Binder, J.X., and Jensen, L.J. (2015). DISEASES: text mining and data integration of disease-gene associations. Methods *74*, 83–89.

Polouliakh, N., Nock, R., Nielsen, F., and Kitano, H. (2009). G-protein coupled receptor signaling architecture of mammalian immune cells. PLoS ONE *4*, e4189.

Ravasz, E., Somera, A.L., Mongru, D.A., Oltvai, Z.N., and Barabási, A.-L. (2002). Hierarchical organization of modularity in metabolic networks. Science *297*, 1551–1555.

Reimand, J., Arak, T., Adler, P., Kolberg, L., Reisberg, S., Peterson, H., and Vilo, J. (2016). g:Profiler—a web server for functional interpretation of gene lists (2016 update). Nucleic Acids Res. *44* (*W1*), W83–W89.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47.

Salas, T.R., Petruseva, I., Lavrik, O., and Saintomé, C. (2009). Evidence for direct contact between the RPA3 subunit of the human replication protein A and single-stranded DNA. Nucleic Acids Res. *37*, 38–46.

Smoot, M.E., Ono, K., Ruscheinski, J., Wang, P.-L., and Ideker, T. (2011). Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics *27*, 431–432.

Svendsen, J.M., Smogorzewska, A., Sowa, M.E., O'Connell, B.C., Gygi, S.P., Elledge, S.J., and Harper, J.W. (2009). Mammalian BTBD12/SLX4 assembles a Holliday junction resolvase and is required for DNA repair. Cell *138*, 63–77.

Tieri, P., Grignolio, A., Zaikin, A., Mishto, M., Remondini, D., Castellani, G.C., and Franceschi, C. (2010). Network, degeneracy and bow tie. Integrating paradigms and architectures to grasp the complexity of the immune system. Theor. Biol. Med. Model. *7*, 32.

Yao, T., Song, L., Jin, J., Cai, Y., Takahashi, H., Swanson, S.K., Washburn, M.P., Florens, L., Conaway, R.C., Cohen, R.E., and Conaway, J.W. (2008). Distinct modes of regulation of the Uch37 deubiquitinating enzyme in the

proteasome and in the Ino80 chromatin-remodeling complex. Mol. Cell *31*, 909–917.

Yeger-Lotem, E., and Sharan, R. (2015). Human protein interaction networks across tissues and diseases. Front. Genet. *6*, 257.

Zediak, V.P., and Berger, S.L. (2008). Hit and run: transient deubiquitylase activity in a chromatin-remodeling complex. Mol. Cell *31*, 773–774.

Zhang, B., and Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. Stat. Appl. Genet. Mol. Biol. *4*, Article 17.

Zhao, B., Hu, S., Li, X., Zhang, F., Tian, Q., and Ni, W. (2016). An efficient method for protein function annotation based on multilayer protein networks. Hum. Genomics *10*, 33.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
BV510 anti-mouse CD45, clone 30-F11	BD Bioscience	563891; RRID: AB_2734134
AF700 anti-mouse CD3e, clone 17A2	BioLegend	100215; RRID: AB_493696
AF700 anti-mouse CD19, clone 6D5	BioLegend	115527; RRID: AB_493734
AF700 anti-mouse B220, clone RA3-6B2	BioLegend	103232; RRID: AB_493717
AF700 anti-mouse TER119, clone TER-119	BioLegend	116218; RRID: AB_528961
AF700 anti-mouse NK-1.1, clone PK136	BioLegend	108729; RRID: AB_2074426
BV711 anti-mouse IA/IE (MHC II), clone M5/114.15.2	BD Bioscience	107643; RRID: AB_2738191
BV605 anti-mouse CD11b, clone M1/70	BioLegend	101237; RRID: AB_11126744
PE-Cy7 anti-mouse CD11c, clone N418	eBioscience	25-0114-82; RRID: AB_469590
BV421 anti-mouse CD103, clone M290	BD Bioscience	562771; RRID: AB_2737783
AF647 anti-mouse CD64, clone X54-5/7.1	BD Bioscience	558539; RRID: AB_647120
Biotin anti-mouse CD3e, clone 17A2	BioLegend	100244; RRID: AB_2563947
Biotin anti-mouse CD19, clone RM4-5	BioLegend	115503; RRID: AB_313638
FITC anti-mouse F4/80, clone BM8	BioLegend	123107; RRID: AB_893500
PE/Dazzle 594 anti-mouse CD172a (SIRPα), clone P84	BioLegend	144015; RRID: AB_2565279
PE anti-mouse XCR1, clone ZET1	BioLegend	148204; RRID: AB_2563843
Chemicals, Peptides, and Recombinant Proteins		
Fetal Calf Serum (FCS)	Sigma Aldrich	F7524
PBS	GIBCO/LifeTechnologies	14190-094
RPMI1640	GIBCO/LifeTechnologies	21875-034
Sodium Pyruvate	GIBCO/LifeTechnologies	11360-039
HEPES	GIBCO/LifeTechnologies	15630-056
Penicillin/Streptomycin solution (100x)	GIBCO/LifeTechnologies	15140-122
Collagenase IV (0.5mg/ml)	Sigma Aldrich	C5138
DNase I (12.5mg/ml)	Sigma Aldrich	D4263
LIVE/DEAD Fixable Near-IR Dead Cell Stain Kit, for 633 or 635 nm excitation	Invitrogen, ThermoFisher Scientific	L10119
Critical Commercial Assays		
EasySep Mouse Streptavidin RapidSpheres	StemCell Technologies, Inc.	19860A
Deposited Data		
RNA sequencing of murine conventional dendritic cell lineage 1 (cDC1) cells from mesenteric lymph nodes	GEO	GSE150549
RNA-seq data of six immune cells (homo sapiens)	ArrayExpress	E-GEOD-60424
Experimental Models: Organisms/Strains		
C57BL6 mice	Janvier	C57BL6
Software and Algorithms		
Bow-tie R package	https://github.com/k-niss/bowtie	bowtie
R	R project	Version 3.6.1
FastQC	Babraham Bioinformatics	Version 0.11.8
STAR RNA-seq aligner	https://github.com/alexdobin/STAR	Version 2.4.2
Cytoscape	Cytoscape.org	Version 3.6.1

RESOURCE AVAILABILITY

Lead contact

Further information and requests for software and data should be directed to and will be fulfilled by the Lead Contact, Søren Brunak (soren.brunak@cpr.ku.dk).

Materials Availability

This study did not create new unique reagents.

Data and Code Availability

The accession number for the raw and processed RNA-seq data of cDC1 cells from murine mesenteric lymph nodes reported in this paper is GEO: GSE150549. Software used to generate wTO matrices, locate protein complexes and locate bow-tie motifs is available as an R package at https://github.com/k-niss/bowtie.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Female 7-10 weeks old C57BL6 mice were purchased from Janvier and maintained at the Biomedical Center (BMC), Lund University. Animal experiments were performed in accordance with the Lund/Malmö Animal Ethics Committee.

METHOD DETAILS

Cell preparation

Cell suspensions of murine MLN were prepared as previously described (Luda et al., 2016). Briefly, MLN were cut into small pieces and enzymatically digested with Collagenase IV (0.5 mg/ml, Sigma-Aldrich) and DNase I (12.5 mg/ml, Sigma Aldrich) for 45 min at 37°C while shaking and filtered prior to analysis. To pre-enrich the samples prior to cell sorting, B and T cells were labeled with biotinylated antibodies to CD19 and CD3, respectively, and depleted by using EasySep Streptavidin RapidSpheres (STEMCELL Technologies) according to manufacturer's instructions.

Flow cytometry analysis and cell sorting

MLN cell suspensions were stained with a cocktail of antibodies to CD45 (30-F11), CD3 (17A2), CD19 (6D5), B220 (RA3-6B2), TER-119 (TER-119), NK-1.1 (PK136), MHC II (IA/I-E) (M5/114.15.2), CD11b (M1/70), CD11c (N418), CD103 (M290), CD64 (X54-5/7.1), F4/80 (BM8), SIRPα (P84), XCR1 (ZET). $CD11c^{+}MHCII^{hi}XCR\text{-}1^{+}CD103^{+}$ cDC1 were flow cytometry cell sorted on a FACSAriaII (BD Biosciences) after exclusion of dead cells stained by fixable live/dead near-infrared dye (Life technologies) and cell aggregates (identified on FSC-A versus FSC-W scatterplots).

Library preparation

Total RNA was isolated using the RNeasy Micro kit (QIAGEN). Briefly, frozen pellets of sorted cells were resuspended in RLT buffer. Buffer and cells were mixed for 30 s through pipetting and left on the bench for 5 min to assure proper lysis. Extraction was performed according to manufacturer's protocol with an on-column DNase digestion step added after the first wash buffer step. RNA quality and quantity were measured using the 2100 BioAnalyzer equipped with RNA6000 Pico chip (Agilent Technologies). RNA was subjected to whole transcriptome amplification using Ovation RNA-Seq System V2 (NuGEN) kit, following the manufacturer's protocol. This kit employs a single primer isothermal amplification (SPIA) method to amplify RNA into double stranded cDNA. Amplified cDNA samples were purified using the MinElute Reaction Cleanup kit (QIAGEN) according to manufacturer's instructions. Quantity and quality of the cDNA samples were measured using the 2100 BioAnalyzer equipped with DNA1000 chip (Agilent technologies) and the Nanodrop (ThermoFisher Scientific). Libraries to sequence with Illumina platform were constructed the Ovation Ultralow system V2 kit (NuGEN), following manufacturer's instructions. 100ng of amplified cDNA were fragmented by sonication using a Bioruptor Pico (Diagenode). The sheared cDNA was end-repaired to generate blunt ends, then ligated to Illumina adaptors with indexing tags, followed by AMPure XP beads purification. The final size distribution of the libraries was evaluated using 2100 Bioanalyzer equipped with DNA1000 chip (Agilent technologies) and quantified using KAPA library Quantification Kit Illumina platforms (Kapa Biosystems).

mRNA sequencing using Illumina

Equimolar amounts of each sample library (n = 9) were pooled and the pools were used for cluster generation on the cBot with the Illumina TruSeq SR Cluster Kit (v3). Sequencing was performed on an Illumina HiSeq 1000 instrument using indexed 50 cycles single-read protocol and the TruSeq SBS v3 Reagents according to the Illumina HiSeq 1000 System User Guide. Image analysis and base calling resulted in BCL files, which were converted to FASTQ files with the CASAVA1.8.2 software. The resulting libraries had a length of 50 base pairs and 25-40 million reads per sample.

RNA-seq data analysis of cDC1 data

The quality of the FASTQ files was initially checked using FastQC (Andrews, 2010). The reads were aligned to the ENSEMBL mouse genome (GRCm38 release 89) using STAR version 2.4.2a (Dobin et al., 2013) with default settings, which also quantified the gene expression. Gene expression files were preprocessed as recommended in Law et al. (Law et al., 2014). Briefly, genes with a log2 count per million (log2-cpm) of at least 1.0 in three out of nine samples were considered expressed. Raw counts were hereafter transformed using the *voom* function of the *limma* R package (Ritchie et al., 2015). The pipeline produced a final table of 10,563 expressed genes across nine samples.

Analysis of sorted immune cell RNA-seq data

FASTQ files associated with the public dataset ArrayExpress: E-GEOD-60424 (Linsley et al., 2014) were downloaded using the fastq-dump function from the sratoolkit (v. 2.8.2). High quality sequencing was confirmed using FastQC (Andrews, 2010). Alignment of the FASTQ files and gene expression quantification was performed as described above in RNA-seq data analysis of cDC1 data. We only used the samples that were metadata tagged as "normal," i.e., from healthy patients.

The cDC1-specific protein interactome

To construct the cDC1-specific protein interactome, we used the protein-protein interactome InWeb_IM (Li et al., 2017) as foundation. InWeb_IM contains experimentally generated and scored physical PPIs from mainly *H. sapiens* (n = 332,385), *m. musculus* (n = 169,347) and *S. cerevisiae* (n = 102,150) (PPI counts from Li et al., 2017). Each PPI is benchmarked against a gold standard and scored accordingly, summed into a confidence score (cs). Only proteins expressed (defined above) in cDC1 cells were kept to generate the cDC1-specifc interactome. This method has previously been suggested in Magger et al. (2012) and applied in Pedersen et al. (2017). Since InWeb_IM is indexed using UniProtKB/Swiss-Prot identifiers of *H. sapiens*, the 10,563 ENSMUSG identifiers of *m. musculus*, which represented the expressed genes in cDC1, were first mapped to *H. sapiens* ENSG identifiers (n = 9,183) using ortholog pairing and then further mapped to *H. sapiens* UniProt/Swiss-Prot identifiers (n = 8,984). The mapping was done using the R package *gProfileR* (Reimand et al., 2016). The final cDC1-specific interactome contained 8,568 expressed proteins, which had at least one PPI in the InWeb_IM database.

The cDC1-specific interactome topology

To describe the topology of the cDC1-specific interactome, we calculated the pairwise wTO between all protein pairs. The wTO has previously been used in an unweighted form in Ravasz et al. (Ravasz et al., 2002) and been refined to its weighted form in Zhang et al. (Zhang and Horvath, 2005). Given a protein x and a protein y in the interactome and with a symbolizing a connection strength between two proteins $(0 \leq a \leq 1)$, the wTO was calculated using:

$$wTO_{xy} = \frac{l_{xy}}{\min\{k_x, k_y\} + 1 - a_{xy}}$$

where $l_{xy} = \sum_u a_{xu} a_{yu}$ is the product of the connection strengths between x with u and y with u, where u is a common protein neighbor of both x and y, summed over all common protein neighbors of x and y. By neighbors, we refer to first order interactors of x and y only. The term $k_x = \sum_u k_{xu}$ is the summed connection strength of protein x to all of its neighbors. For each protein pair x and y, this will produce a wTO between 0 and 1. Since we based our analysis on InWeb_IM, we used InWeb's internal confidence score $(0 \leq cs \leq 1)$ as a measure of inter-protein connection strength. The all proteins versus all proteins wTO matrix was constructed, resulting in an 8,568 × 8,568 symmetric wTO matrix (36.7 million unique pairwise comparisons).

Matrix and network visualization

We produced a visualization of the cDC1 interactome topology in R by clustering the cDC1 wTO matrix via *FlashClust* using the unweighted pair group method with arithmetic mean (UPGMA) (Langfelder, 2012), which resulted in a meaningful protein order, as previously demonstrated by Ravasz et al. (Ravasz et al., 2002). We visualized the resulting matrix using the basic R function *image* with raster activated, creating a 73.4-megapixel image. All graph visualizations were done in Cytoscape (Smoot et al., 2011).

Bow-tie motif localization in wTO matrix

We used an adjacency matrix of the cDC1 interactome, where each pairwise PPI were represented by its InWeb_IM confidence score. A score of zero indicated no interaction. The adjacency matrix was sorted row- and column-wise using the protein order of the hierarchically clustered cDC1 wTO matrix (see *Matrix and network visualization*). The interaction fans of bow-tie motifs were observed as line substructures situated between two square substructures, i.e., protein complexes, in the wTO matrix. To locate all interaction fans, we iterated over each row in the adjacency matrix, searching for strings of interactions that had a length of > = 10 and where the average confidence score was > = 0.9. We filtered away the strings that represented protein complexes (see *Protein complex localization in wTO matrix*) and kept only the strings of interactions that were located between two protein complexes.

Protein complex localization in wTO matrix

We used the adjacency matrix described above in localization of bow-tie motifs in the cDC1 interactome matrix. To find protein complexes which were represented by square substructures in the wTO matrix, we searched for quadrilateral square submatrices on the diagonal of the adjacency matrix that had a high average confidence score. We used each protein as a "seed." Around each protein seed's position on the diagonal of the adjacency matrix, we first isolated a 3x3 submatrix with the seed in the middle. If the submatrix had an average confidence score threshold of > 0.75, we expanded the submatrix by 1 protein in each direction, producing a 5x5 submatrix. If that submatrix also passed the 0.75 threshold, we again expanded by +1 in each direction and continued until (i) the average confidence score got below the threshold or (ii) we reached a 500x500 submatrix. The position of the largest submatrix for each seed protein was saved. Overlapping submatrices were merged and a minimum size of 9x9 proteins was enforced.

We chose an average confidence score threshold of > 0.75, as we found by visual inspection that this value was good at highlighting the dark submatrices on the diagonal in the PPI adjacency matrix (See red highlights in supplementary Figure S2). Furthermore, a stability test of average confidence scores showed that thresholds values between 0.65 to 0.85 produced a largely similar number of protein complexes (Figure S1).

DISEASES and COMPARTMENTS databases

To obtain protein sets related to diseases, we downloaded the filtered knowledge channel from the DISEASE database (https://diseases.jensenlab.org/downloads) on April 2019. We only used proteins who had a confidence score of four or more. The ENSP identifiers used in the DISEASE database were converted into UNIPROT identifiers using the *gconvert* function of gProfiler. To obtain the subcellular location of each protein, we downloaded the data of the human integrated channel from the COMPARMENTS database (http://compartments.jensenlab.org/downloads) on April 2019. We again only used proteins with a confidence score of four or more. To limit redundancy, we only used the following compartment terms: Mitochondrion, Golgi apparatus, Endoplasmic reticulum, Cytosol, Lysosome, Endosome, Peroxisome, Plasma membrane, Cytoskeleton, Nucleus, Extracellular region.

Membraneless organelles

For enrichment permutation test of "knot" proteins in membraneless organelles, we tested for a significant overlap between the "knot" protein set and ten different protein sets of membraneless organelles, as defined by their gene ontology term: Nucleolus (GO:0005730), Cajal body (GO:0015030), Nuclear speck (GO:0016607), Paraspeckle (GO:0042382), Histone locus body (GO:0035363), PML body (GO:0016605), P body (GO:0000932), Cytoplasmic stress granule (GO:0010494), Nuclear stress granule (GO:0097165) and P granule (GO:0043186). To test for enrichment of proteins belonging to membraneless organelles in our protein complexes, we performed a gene ontology test with gProfiler (Reimand et al., 2016) testing only for enrichment of cellular component GO terms and using the cDC1 protein interactome as custom background. We hereafter checked if any of the above-listed ten GO terms were found enriched.

Protein mass, domains and isoforms

Information on protein mass, domains and isoforms were obtained by querying all 8,568 proteins of the cDC1-specifc protein interactome via uniprot.org, exporting the data table and extracting information from the Alternative products (isoforms), Domain and Mass columns. Data was obtained February 2019.

QUANTIFICATION AND STATISTICAL ANALYSIS

For the permutation tests, we compared the specified value (mean protein mass, number of associated gene ontologies, etc.) of the "knot" protein set to 10,000 other random protein sets of the same size (n = 294). The random protein sets were either sampled using a random approach without replacement or by a random degree-adjusted approach without replacement. For the degree-adjusted approach, we assigned all of the 8,568 proteins in the cDC1-specific interactome into 40 different bins depending on the number of PPIs that each protein had, i.e., their degree. This produced 40 bins with about 215 proteins in each. To create a random degree-adjusted sample of 294 proteins, we picked out random proteins from the bins that corresponded to the ones wherein the "knot" proteins were assigned. We hereby obtained random protein sets with a degree distribution similar to that of the "knot" proteins. An enrichment *P value* was calculated by summing the number of times a random protein set out of 10,000 had a value higher than or equal to that of the "knot" protein set. This sum was divided by 10,000 to generate a *P value*. For the underrepresentation *P value*, the number of times a random protein set value was lower than or equal to that of the "knot" protein set value was summed.